

Table S4: Results of the RepeatScout program

Genome	k-mers	Consensus	Mean length (median)	S _n *	S _p *	R _{cc}
<i>D. mel.</i>	5,624,530	1,770	552 (209)	94.87%	57.63%	25.00%
<i>A. tha.</i>	7,569,031	3,417	498 (221)	82.95%	39.54%	13.00%

S_n*: percentage of “knowledge-based” consensus sequences matching a *de novo* consensus sequence

S_p*: percentage of *de novo* consensus sequences matching a “knowledge-based” consensus sequence

R_{cc}: percentage of fully recovered “knowledge-based” consensus sequences